Decoding in Neural Systems: Stimulus Reconstruction from Nonlinear Encoding

Garrett B. Stanley & Alireza SeyedBoloori
Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA.

Abstract— **The encoding of information about the outside world in the temporal activity of sensory neurons is an extremely complex process that has eluded the understanding of the scientific community for decades. The reconstruction of sensory stimuli from observed neuronal activity provides a basis within which we might ascertain the nature of the sensory information encoded by the cells. We present a decoding strategy for predicting the sensory stimulus from the neuronal response that is based on the mechanisms of encoding. For a class of encoding mechanisms characterized by a linear function followed by a memoryless nonlinearity, referred to as Wiener systems, the Bayesian estimator is derived from the transformational properties of the nonlinearity. The result is a reconstruction paradigm in which the ability to predict sensory stimuli from the neuronal response depends heavily upon how well the encoding process has been characterized, and thus provides a measure of our understanding of the underlying physiological process.**

Keywords— **neural systems, decoding, estimation**

I. Introduction

The reconstruction of sensory inputs from recorded neural activity has proved to be an invaluable tool in understanding how the information about the outside world is encoded in the sensory pathway. We and others have shown previously that a surprising amount of detail can be reconstructed from ensemble neural activity in sensory pathways using a relatively simple linear reconstruction technique [1], [2], [3]. The accepted approach involves the correlation of the neural response with the sensory stimulus in order to determine the optimal "reverse-filter" that predicts the sensory input from the neuronal activity.

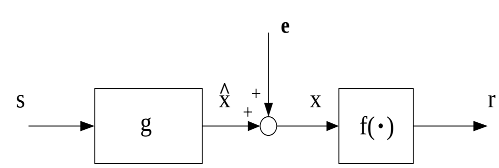
Although this approach has been useful in evaluating what and how much information is being encoded in the neural activity, it does not directly utilize the knowledge of the underlying encoding mechanisms and therefore provides little or no insight in this regard. In fact, the dominant belief is that the decoding mechanism can remain linear even if the encoding process is highly nonlinear in nature, although there has been no rigorous support for this claim. Recent work in place cell encoding in the hippocampus has begun to shed some light on the utility of the encoding/decoding approach [4], [5], but this has yet to become the mainstream perspective.

Due to the inherent nonlinearity of encoding continuous sensory information with a discrete neuronal process, no neural system is purely linear in its form. Obviously general, non-parametric nonlinear models could be used to represent the encoding process. However, many early sensory systems are well characterized as a linear system followed by a static, memoryless nonlinearity. This type of cascade system has historically been referred to as a Wiener system or LN system, and is structurally much more simple than the corresponding general Wiener series expansion that would be necessary to characterize the same dynamics. The result is that the output is representative of the neuronal firing rate, rather than the discrete events that are observed experimentally. A common perspective is that the output of the Wiener system is the rate of an inhomogeneous Poisson process that results in the stochasticity of the observed event times. Given such an encoding mechanism, the question then remains as to how we might decode the sensory input from the observed neuronal activity, and what limitations might the encoding process place on the prediction.

II. The Wiener System

The Wiener system is a class of nonlinear systems in which a linear dynamical system is followed by a static, memoryless nonlinearity, as shown in Figure 1. The Wiener system has been widely utilized to describe

Fig. 1. Wiener System



the relationship between the stimulus and the firing rate of a neuron [6], [7], [8]. In contrast to the complex nature of a higher order Wiener kernel representation, the Wiener system provides a relatively simple means for describing the inherent nonlinearity in neural encoding. A simple half-wave rectification achieves the non-negative characteristics of the firing rate, and has also been widely used as the static nonlinearity in the

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Decoding in Neural Systems: Stimulus Reconstruction From Nonlinear Encoding	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Division of Engineering and Applied Sciences Harvard University Cambridge, MA 02138	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

encoding process.

We can express the dynamics of the system shown in Figure 1 as $\mathbf{r} = f(\mathbf{g} * \mathbf{s} + \mathbf{e})$, where $\mathbf{r} \in \mathbb{R}^{N\times 1}$ is the firing rate of the neuron over N time steps, $\mathbf{s} \in \mathbb{R}^{N\times 1}$ is the corresponding sensory stimulus, $\mathbf{g} \in \mathbb{R}^{L\times 1}$ is the first order Wiener kernel, $\mathbf{e} \in \mathbb{R}^{N\times 1}$ is a white Gaussian noise process, $f(\cdot)$ is the static nonlinearity, and $*$ represents convolution.

It has been shown previously that the linear block of the system can be estimated as $\hat{\mathbf{g}} = a\Phi_{ss}^{-1}\phi_{sr}$, where $\Phi_{ss} \in \mathbb{R}^{L\times L}$ is the input auto-covariance Toeplitz matrix, $\phi_{sr} \in \mathbb{R}^{L\times 1}$ is the cross- covariance between the input and the output, and $a \in \mathbb{R}$ is a scaling factor [9]. The scaling factor is a result of the static non-linearity, and is identically 2 when the non-linearity is a half-wave rectification (Stanley, unpublished).

III. Decoding from the Wiener System

Current approaches to decoding are based on techniques that utilize the correlation between the stimulus and response, normalized by the correlation structure in the response. In this paper, we will refer to this approach as the "reverse filter" approach, since the output is essentially treated as an input, and vice versa. Although this perspective provides a tool for reconstructing sensory inputs from recorded neural activity [2], [3], it makes no direct reference to the underlying encoding mechanisms.

Alternatively, knowledge of the encoding strategy can instead lead us to questions concerning the decoding of sensory stimuli from neural activity based on the nature of the encoding process. The decoding problem involves the estimation of the input $\mathbf{s}$ from the recorded neuronal response, *given* the encoding strategy. For the Wiener systems described above, this assumes prior identification of the linear stage utilizing an independent data set, although the basic premise described here applies to more general model structures. Within a Bayesian framework, we pose this problem as:

$$\hat{\mathbf{s}}_{MMSE} = E\{\mathbf{s}|\mathbf{r}\} \qquad (1)$$

The optimal estimator, that minimizes the Bayesian mean-square error (MMSE), is the expected value of the stimulus conditioned on the response. Embedded in the conditional density is the underlying encoding mechanism, determined independently. For jointly Gaussian processes, the problem reduces to a well known relationship. In the problem presented here, we will restrict the input $\mathbf{s}$ and noise $\mathbf{e}$ to Gaussian white processes, but $\mathbf{s}$ and $\mathbf{r}$ will not be jointly Gaussian due to the effect of the static nonlinearity. We can, however, determine the conditional density $p(\mathbf{s}|\mathbf{r})$. In the general case, in order to determine $E\{\mathbf{s}|\mathbf{r}\}$, we must perform the following integration:

$$E\{\mathbf{s}|\mathbf{r}\} = \int_{-\infty}^{\infty} \mathbf{s}p(\mathbf{s}|\mathbf{r})d\mathbf{s} \qquad (2)$$

where $p(\mathbf{s}|\mathbf{r})$ is the conditional density. We can write the conditional density as:

$$p(\mathbf{s}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s})p(\mathbf{s})}{p(\mathbf{r})} \qquad (3)$$

Given the stimulus $\mathbf{s}$, the responses at different times are independent, giving:

$$p(\mathbf{r}|\mathbf{s}) = \prod_k p(r[k]|\mathbf{s}) \qquad (4)$$

As shown in the appendix, for half-wave rectification we can write:

$$\begin{aligned} p(r[k]|\mathbf{s}) &= \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{(r[k]-\hat{x}[k])^2}{2\sigma_e^2}} U(r[k]) \\ &\quad + \frac{\delta(r[k])}{\sqrt{2\pi\sigma_e^2}} \int_{-\infty}^{0} e^{-\frac{(x-\hat{x}[k])^2}{2\sigma_e^2}} dx \end{aligned} \qquad (5)$$

where:

$$\hat{x}[k] = \sum_{m=0}^{L-1} g[m]s[k-m]$$

$U(r[k])$ is the unit step function, resulting in 1 for $r[k] > 0$, zero else, and $\delta(\cdot)$ is the Dirac delta function. The density of the stimulus is $p(\mathbf{s}) \sim N(0, \sigma_s^2 I)$, completing the numerator of Equation 3. The denominator is simply a scaling term, and can be obtained from the integration of the numerator over $\mathbf{s}$. The Bayesian estimator can then be obtained from the integration in Equation 2. However, the estimation then requires an N-dimensional integration, which is impractical for any non-trivial data length.

An alternative to the MMSE estimator is the maximum *a posteriori* (MAP) estimator, which instead involves a maximization of the conditional density:

$$\mathbf{s}_{MAP} = \arg\max_{\mathbf{s}} p(\mathbf{s}|\mathbf{r}) = \arg\max_{\mathbf{s}} p(\mathbf{r}|\mathbf{s})p(\mathbf{s}) \qquad (6)$$

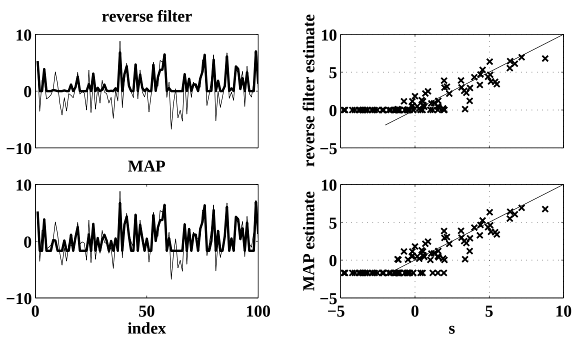
which now is an N-dimensional optimization problem, and therefore becomes tractable for relatively small data sets, whereas the corresponding N-dimensional integration is not feasible. For larger N, the data sets can be segmented for computational efficiency. Numerical searches were implemented using a gradient search method (quasi-newton).

So we now have two alternate methods by which we may reconstruct, or decode, sensory inputs from recorded neuronal responses: the reverse filter approach and the Bayesian MAP approach. In subsequent analyses, we will compare the two techniques.

IV. Results

First consider a static case, where the linear system **g** is a scalar value, or equivalently proportional to an impulse function. For the example shown here, the gain of the static linearity is positive. The output of the linear stage is then passed through a half-wave rectification, giving $r = f(gs+e)$. Figure 2 shows the results from this simple case. We can see that for the case when

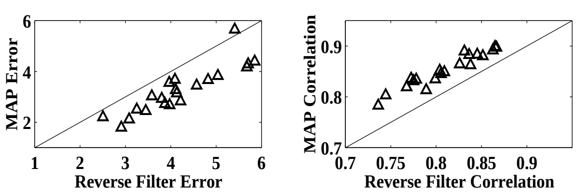
Fig. 2. Wiener System with Static Linearity



Results from reverse filter and MAP estimator for static Wiener system. The left two panels show the actual (thin) and reconstructed (thick) stimuli for the reverse filter approach (top) and the MAP approach (bottom). Scatter plots are shown on the right of actual stimulus value versus the predicted value from the two techniques.

the stimulus is positive, the two estimates are nearly identical. The difference in the two estimates is evident when the stimulus is negative. The reverse filter approach provides an estimate of zero, while the MAP estimator yields a value that is slightly negative. Running many such examples revealed that the MAP estimator provided a significantly better prediction of the stimulus than that obtained from the reverse filter, as shown in Figure 3. In these plots, data points above

Fig. 3. Summary Statistics for Static Case

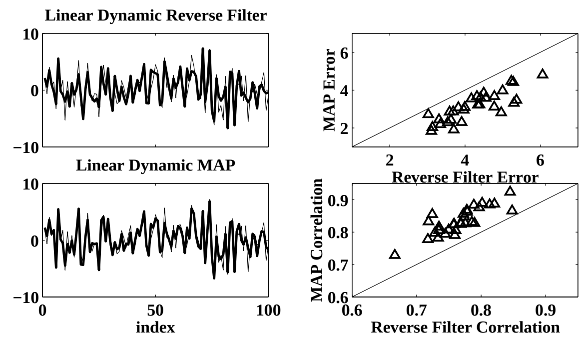


Left plot shows the mean square error of the reconstruction for the reverse filter approach versus that for the MAP approach for the static Wiener system for several different simulations. Right plot shows the same for the correlation between actual and estimated stimulus.

the line indicate that the measure associated with the MAP estimator is greater and vice versa. So, for example, in Figure 3, the MAP estimator provides smaller prediction error and larger correlation.

With the dynamic case, we first consider an entirely linear system, where $\mathbf{r} = \mathbf{g} * \mathbf{s} + \mathbf{e}$. Again, we see that the MAP estimator provides significantly better results than those obtained from the reverse filter method, as shown in Figure 4. For the linear case, we were able

Fig. 4. Linear Dynamical System

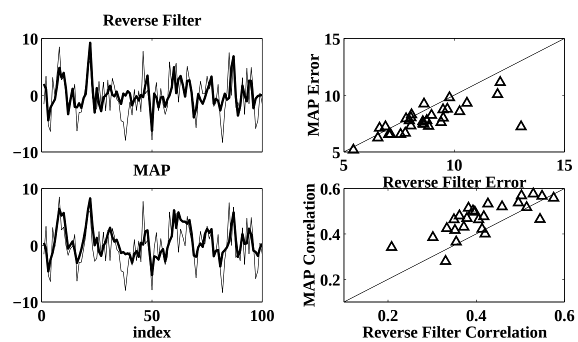


Results from reverse filter and MAP estimator for dynamic linear system. The left two panels show the actual (thin) and reconstructed (thick) stimuli for the reverse filter approach (top) and the MAP approach (bottom). The right plots show the corresponding scatter plots of estimation error (top) and correlation (bottom) for the reverse filter estimate versus the MAP estimate.

to compute both the MMSE estimator and the MAP estimator for comparison, and found that they were not significantly different.

Finally, the results from the dynamic Wiener system are shown in Figure 5. In this case, the response is the

Fig. 5. Dynamical Wiener System



Results from reverse filter and MAP estimator for dynamic Wiener system. The two panels on the left show the actual (thin) and reconstructed (thick) stimuli for the reverse filter approach (top) and the MAP approach (middle). The right plots show the corresponding scatter plots of estimation error (top) and correlation (bottom) for the reverse filter estimate versus the MAP estimate.

half-wave rectified output of the linear system, and is written $\mathbf{r} = f(\mathbf{g} * \mathbf{s} + \mathbf{e})$. Again we see that the MAP approach tends to outperform the reverse filter approach, although the effect is not so dramatic here. As a confirmation that the gradient search method was indeed providing a more optimal solution in the MAP sense, the cost function given in 6 was evaluated at both the MAP estimate for **s** and the reverse filter estimate for **s** (data not shown). The MAP estimate always provided a significantly larger $p(\mathbf{r}|\mathbf{s})$.

V. DISCUSSION

We have shown that knowledge of the mechanisms by which neurons encode sensory information can significantly enhance the process of decoding the sensory stimulus from the recorded neuronal activity. The reverse filter approach, in which the correlation structure between response and stimulus is used to predict the stimulus from the neuronal response, assumes independence between the response and the stimulus prediction error, although this is not generally the case.

The Bayesian estimator, as an alternate approach, is based on the underlying conditional density functions between stimulus and response. The difference between the reverse filter approach and the Bayesian approach is perhaps most clearly outlined with the static case with the half-wave rectification discussed previously. The reverse filter approach provides a clean estimate of the corresponding stimulus for positive values of response, but when the response is rectified to zero, the predicted stimulus is also zero. The MAP estimator, however, provides a negative offset for a rectified response, essentially relying on the fact that we know something about the distribution of the stimulus even when the response is rectified, and yields the best estimate in the sense of maximizing the posterier density. The same heuristic argument holds for the dynamic case, providing us with an improved reconstruction of the sensory stimulus from the recorded response.

The methodology presented here holds for more general types of nonlinearities, although the transformation of density functions is case dependent. The assumption of a Gaussian white noise stimulus can be relaxed for a more general treatment of the problem, although the assumption of additive Gaussian white noise is critical for our current derivation. The result of this work is a step towards recovery of information loss due to the inherent rectifying properties in neuronal encoding.

ACKNOWLEDGMENTS

We thank Emery Brown for helpful comments regarding the density transformations.

APPENDIX

Consider a random variable r that is a nonlinear function $f(\cdot)$ of a random variable x, which is the sum of random variable s scaled by a known parameter g with an additive noise term e, given by $r = f(x) = f(gs+e)$. The scaling $g \in \mathbb{R}$ is assumed known, as are the densities $p_s(s)$ and $p_e(e)$. The stimulus s is assumed zero-mean Gaussian $\sim N(0,\sigma_s^2)$, as is the noise $e \sim N(0,\sigma_e^2)$. The output r is observed. The best estimate of s, given the observation of r, is:

$$\hat{s} \quad = \quad E\{s|r\} = \int_{-\infty}^{\infty} s p(s|r) ds \qquad (7)$$

where $p(s|r)$ is the conditional density of the stimulus given the response. The conditional density can be expressed as:

$$p(s|r) \quad = \quad \frac{p(r|s)p(s)}{p(r)} = \frac{p(r|s)p(s)}{\int p(r|s)p(s)ds}$$

Given the stimulus s, the density of x can be written as:

$$p_{x|s}(x) \quad = \quad \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{-\frac{(x-gs)^2}{2\sigma_e^2}\right\}$$

If the static nonlinearity is a half-wave rectification, then the density of the response given the stimulus becomes [10], [11]:

$$\begin{aligned} p(r|s) &= p_x(r)U(r) + P_x(0)\delta(r) \\ &= \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{-\frac{(r-gs)^2}{2\sigma_e^2}\right\} U(r) \\ &\quad + \delta(r) \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{-\frac{(x-gs)^2}{2\sigma_e^2}\right\} dx \end{aligned}$$

where $U(r)$ is the unit step function, which is 1 for $r > 0$, and 0 else, and $P_x(x)$ is the cumulative distribution function for x. The effect of the half-wave rectification is to concentrate the density for negative values of input to the nonlinearity at 0 in the density of the output of the nonlinearity. The estimate in Equation 7 can then be evaluated numerically using the conditional density $p(r|s)$ described above.

REFERENCES

[1] W. Bialek, F. Rieke, R.R. de Ruyter Van Steveninck, and D. Warland, "Reading a neural code," *Science*, vol. 252, pp. 1854–1857, 1991.

[2] D.K. Warland, P. Reinagel, and M. Meister, "Decoding visual information from a population of retinal ganglion cells," *J. Neurophysiol.*, vol. 78, pp. 2336–2350, 1997.

[3] G.B. Stanley, F.F. Li, and Y. Dan, "Reconstruction of natural scenes from ensemble responses in the lgn," *J. Neurosci.*, vol. 19, no. 18, pp. 8036–8042, 1999.

[4] E.N. Brown, L.M. Frank, D. Tang, M.C. Quirk, and M.A. Wilson, "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells," *J. Neurosci.*, vol. 18, no. 18, pp. 7411–7425, 1998.

[5] K. Zhang, I. Ginzburg, B.L. McNaughton, and T.J. Sejnowski, "Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells," *J. Neurophysiol.*, vol. 79, pp. 1017–1044, 1998.

[6] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, "Spatial summation in the receptive fields of simple cells in the cat's striate cortex," *J. Physiol.*, vol. 283, pp. 53–77, 1978.

[7] D. J. Tolhurst and A. F. Dean, "The effects of contrast on the linearity of the spatial summation of simple cells in the cat's striate cortex," *Experimental Brain Research*, vol. 79, pp. 582–588, 1990.

[8] D. Ringach, G. Sapiro, and R. Shapley, "A subspace reverse-correlation technique for the study of visual neurons," *Vision Res.*, vol. 37, pp. 2455–2464, 1997.

[9] I. W. Hunter and M. J. Korenberg, "The identification of nonlinear biological systems: wiener and hammerstein cascade models," *Biological Cybernetics*, vol. 55, pp. 135–144, 1986.

[10] J. J. Bussgang, "Crosscorrelation functions of amplitude distorted gaussian signals," *MIT Res Lab Elec Tech Rep*, vol. 216, pp. 1–14, 1952.

[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, New York, second edition, 1984.